J. Appl. Prob. **55,** 571–592 (2018)
*doi:*10.1017/jpr.2018.36

COMPUTABLE APPROXIMATIONS FOR AVERAGE MARKOV DECISION PROCESSES IN CONTINUOUS TIME

JONATHA ANSELMI,* *INRIA*
FRANÇOIS DUFOUR,** *INRIA and Université de Bordeaux*
TOMÁS PRIETO-RUMEAU,*** *UED*

Abstract

In this paper we study the numerical approximation of the optimal long-run average cost of a continuous-time Markov decision process, with Borel state and action spaces, and with bounded transition and reward rates. Our approach uses a suitable discretization of the state and action spaces to approximate the original control model. The approximation error for the optimal average reward is then bounded by a linear combination of coefficients related to the discretization of the state and action spaces, namely, the Wasserstein distance between an underlying probability measure μ and a measure with finite support, and the Hausdorff distance between the original and the discretized actions sets. When approximating μ with its empirical probability measure we obtain convergence in probability at an exponential rate. An application to a queueing system is presented.

Keywords: Continuous-time Markov decision process; Lipschitz continuous control model; approximation of the optimal value function

2010 Mathematics Subject Classification: Primary 90C40
Secondary 90C39

1. Introduction

In this paper we are concerned with numerical methods for approximating the solution of continuous-time Markov decision processes (CTMDPs). Namely, we are interested in approximating numerically the value function of a general Markov decision process (MDP) with Borel state and action spaces, under the expected long-run average cost optimality criterion.

From a theoretical point of view, CTMDPs have been extensively studied. There exist two main techniques to analyze such optimization problems: the dynamic programming and the linear programming approaches. While these two methods are known to be very efficient for establishing different mathematical properties (such as the existence of optimal policies, smoothness of the value function, sufficiency of subclasses of particular policies, and so on), the problem of solving explicitly or numerically a CTMDP remains a critical issue. Indeed, except for a very few specific models, the determination of an optimal policy and the value function is an extremely difficult problem to tackle. In this context, the standard approach

Received 13 June 2017; revision received 31 January 2018.
* Postal address: INRIA Bordeaux Sud Ouest, Bureau B430, 200 av. de la Vieille Tour, 33405 Talence cedex, France. Email address: jonatha.anselmi@inria.fr
** Postal address: Institut de Mathématiques de Bordeaux, Université de Bordeaux, 351 cours de la Libération, F33405 Talence, France. Email address: francois.dufour@math.u-bordeaux.fr
*** Postal address: Statistics Department, UNED, calle Senda del Rey 9, 28040 Madrid, Spain.
Email address: tprieto@ccia.uned.es

for solving an MDP is to develop numerical methods to obtain quasi-optimal solutions. This topic is, therefore, of crucial importance to demonstrate the practical interest of CTMDP as a powerful modeling tool.

In the discrete-time framework, several techniques have been proposed to solve numerically an MDP, and these can be classified into two groups. The first class is dedicated to the study of MDPs with discrete (large finite or countable) state and action spaces. These approaches are mainly related to stochastic approximation techniques such as reinforcement learning, neuro-dynamic programming, approximate dynamic programs, and simulation-based methods; see, e.g. the survey [20] and the books [2], [4], [15], and [19]. The second category is focused on general MDPs with uncountable Borel state and action spaces. For such models, the traditional approach is to approximate the original control problem by means of an MDP with finite state and action spaces whose optimal cost and policies approximate those of the primary control model; see [5]–[8], and [18] and the references therein.

In the continuous-time context, there exist very few results on this topic. In [11], [16], and [17], the authors studied an approximation procedure in which a control model with a denumerable state space and possibly unbounded transition rates was approximated by a sequence of auxiliary control models. Conditions were provided on the sequence of approximating control models ensuring that the corresponding optimal value and optimal policies converged to the value function and the optimal policies of the original model. Such an approach can be found in [16] for unconstrained infinite-horizon discounted CTMDPs in [11] for constrained CTMDPs and in [17] for average reward CTMDPs.

Our objective in this paper is to propose a method for approximating the value function of a continuous-time control model $\mathcal{M}$ with Borel state space $\boldsymbol{X}$ and Borel action space $\boldsymbol{A}$ under the expected long-run average cost optimality criterion. Our approach consists in discretizing the state and action spaces of $\mathcal{M}$ by introducing a new model $\mathcal{M}_{k,\delta}$ as follows.

- *Discretization of the state space.* We assume that the positive part $q^+(\mathrm{d}y \mid x, a)$ of the transition rate $q(\mathrm{d}y \mid x, a)$ governing the dynamics of the control model $\mathcal{M}$ is absolutely continuous with respect to some reference probability measure μ on $\boldsymbol{X}$. We will then replace the state space $\boldsymbol{X}$ with the finite support of a probability measure μ_k (typically, supported on k points) which is an approximation of μ.
- *Discretization of the action sets.* We will replace the action sets $\boldsymbol{A}(x)$ with some finite $\boldsymbol{A}_\delta(x)$ parametrized by $\delta > 0$. The sets $\boldsymbol{A}(x)$ and $\boldsymbol{A}_\delta(x)$ become 'closer' as $\delta \downarrow 0$.

In this context, and after a suitable definition of the control model $\mathcal{M}_{k,\delta}$, we show in Section 3 that the difference between the optimal values $\mathcal{J}^*$ and $\mathcal{J}^*_{k,\delta}$ (of $\mathcal{M}$ and $\mathcal{M}_{k,\delta}$) can be controlled through the Wasserstein distance $\mathcal{W}(\mu, \mu_k)$ between μ and μ_k and the Hausdorff distance between $\boldsymbol{A}(x)$ and $\boldsymbol{A}_\delta(x)$, which is of order δ; namely,

$$|\mathcal{J}^* - \mathcal{J}^*_{k,\delta}| = O(\mathcal{W}(\mu, \mu_k)) + O(\delta).$$

The interesting feature of the discretization procedure that we propose here is that we are able to control explicitly the approximation error and that we obtain *nonasymptotic bounds* depending on $\mathcal{W}(\mu, \mu_k)$ and δ for every $k \geq 1$ and $\delta > 0$. In particular, it is important to mention that we manage to discretize a continuous (not necessarily compact) state space $\boldsymbol{X}$ into a finite set, and that the corresponding discretization error is measured in terms of Wasserstein distance of measures.

Regarding the construction of a probability measure μ_k with finite support that approximates the reference measure μ in the Wasserstein metric, two main approaches exist. One consists in

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

deriving μ_k starting from a covering of $\boldsymbol{X}$ with small radius. This *deterministic* approach allows the distance $\mathcal{W}(\mu, \mu_k)$ to be tightly controlled but it may pose an additional computational challenge. Another possibility is to use a *random* approximation by considering the empirical probability measure μ_k obtained from k independent and identically distributed draws with distribution μ. The approximation error $\mathcal{W}(\mu, \mu_k)$ (which is a random variable) is then measured with a concentration inequality for the nonasymptotic deviation. In this case, the approximation errors converge in probability to 0 at an exponential rate in the sample size k. We will discuss both approaches.

Finally, we would like to mention that the approaches developed in [6]–[8] for the approximation of MDPs in discrete time cannot be used to approximate the value function of the control model $\mathcal{M}$. Indeed, the discrete-time control model obtained when applying the well-known uniformization technique to $\mathcal{M}$ does not satisfy—in general—the absolute continuity condition of [6]–[8]. Further details on this issue are discussed at the end of Section 4.

The rest of the paper is organized as follows. After introducing some notation, we define the control model $\mathcal{M}$ and state our assumptions in Section 2. We show how to approximate the model $\mathcal{M}$ with a finite state and action control model $\mathcal{M}_{k,\delta}$ in Section 3. In Section 4 we study the particular approximations of the reference probability measure μ (both empirical and deterministic). Finally, we present a numerical application in Section 5.

Notation. The set of nonnegative integers is $\mathbb{N} = \{0, 1, 2, \ldots\}$, while the set of real numbers is $\mathbb{R}$. The subscript '$*$' and the superscript '$+$' will refer to the nonzero and nonnegative elements in the corresponding set, respectively. By $\overline{\mathbb{R}}$ we will denote the set of extended real numbers. Combinations of these indices will yield the corresponding sets. The symbols '$\wedge$' and '$\vee$' denote 'minimum' and 'maximum', respectively.

Given a Borel space $\boldsymbol{Y}$ with metric $d_{\boldsymbol{Y}}$, its Borel σ-algebra will be denoted by $\mathcal{B}(\boldsymbol{Y})$. In this paper, measurability is always referred to the Borel σ-algebra. We say that a function $v\colon \boldsymbol{Y} \to \boldsymbol{Z}$, where $\boldsymbol{Y}$ and $\boldsymbol{Z}$ are Borel spaces, is Lipschitz continuous if there exists $L \geq 0$ with

$$d_{\boldsymbol{Z}}(v(x), v(y)) \leq L d_{\boldsymbol{Y}}(x, y) \quad \text{for all } x, y \in \boldsymbol{Y}. \tag{1}$$

In this case, we will say that v is L-Lipschitz continuous.

Let $\mathbb{B}(\boldsymbol{Y})$, $\mathbb{C}(\boldsymbol{Y})$, and $\mathbb{L}(\boldsymbol{Y})$ denote the families of real-valued functions on $\boldsymbol{Y}$ which are bounded and measurable, bounded and continuous, and Lipschitz continuous, respectively. The supremum norm of $v \in \mathbb{B}(\boldsymbol{Y})$ is $\|v\|$. Given a measurable function $h\colon \boldsymbol{Y} \to [1, \infty)$, the family of measurable functions $v\colon \boldsymbol{Y} \to \mathbb{R}$ such that

$$\|v\|_h = \sup_{x \in \boldsymbol{Y}} \left\{ \frac{|v(x)|}{h(x)} \right\} < \infty$$

will be denoted by $\mathbb{B}_h(\boldsymbol{X})$. If, in addition, v is Lipschitz continuous, we will write $v \in \mathbb{L}_h(\boldsymbol{X})$.

We say that $Q\colon \mathcal{B}(\boldsymbol{Y}) \times \boldsymbol{Z} \to \overline{\mathbb{R}}^+$ is a transition measure on the Borel space $\boldsymbol{Y}$ given the Borel space $\boldsymbol{Z}$ if $B \mapsto Q(B \mid z)$ is a (nonnegative) measure on $(\boldsymbol{Y}, \mathcal{B}(\boldsymbol{Y}))$ for all $z \in \boldsymbol{Z}$ and $z \mapsto Q(B \mid z)$ is measurable for every $B \in \mathcal{B}(\boldsymbol{Y})$. For measurable $v\colon \boldsymbol{Y} \to \mathbb{R}$, we will denote by Qv the function on $\boldsymbol{Z}$ defined as

$$Qv(z) = \int_{\boldsymbol{Y}} v(y) Q(\mathrm{d}y \mid z) \quad \text{for } z \in \boldsymbol{Z},$$

whenever the integral is well defined. The indicator function of a set B will be denoted by $\mathbf{1}_B$. On the product $\boldsymbol{Y} \times \boldsymbol{Z}$ of Borel spaces, we will consider the taxicab metric.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

On the family of nonempty closed subsets of a Borel space $\boldsymbol{Z}$, we will consider the Hausdorff metric defined as

$$d_{\mathrm{H}}(C_1, C_2) = \sup_{z_1 \in C_1} \inf_{z_2 \in C_2} \{d_{\boldsymbol{Z}}(z_1, z_2)\} \vee \sup_{z_2 \in C_2} \inf_{z_1 \in C_1} \{d_{\boldsymbol{Z}}(z_1, z_2)\}.$$

It is known that d_{H} is a metric except that it might not be finite. Lipschitz continuity of a closed-valued multifunction ψ from $\boldsymbol{Y}$ to $\boldsymbol{Z}$ is defined as in (1) for the metric d_{H}.

The family of probability measures on $(\boldsymbol{Y}, \mathcal{B}(\boldsymbol{Y}))$ is denoted by $\mathcal{P}(\boldsymbol{Y})$. Given $y \in \boldsymbol{Y}$, the Dirac probability measure at y will be denoted by δ_y. The family of probability measures $\mu \in \mathcal{P}(\boldsymbol{Y})$ with finite first moment (that is, $\int_{\boldsymbol{Y}} d_{\boldsymbol{Y}}(y, y_0)\mu(\mathrm{d}y) < \infty$ for some, and then for all, $y_0 \in \boldsymbol{Y}$) is denoted by $\mathcal{P}_1(\boldsymbol{Y})$. We say that the probability measure $\mu \in \mathcal{P}(\boldsymbol{Y})$ has a finite exponential moment if there exists $\gamma > 0$ such that

$$\int_{\boldsymbol{Y}} \exp\{\gamma d_{\boldsymbol{Y}}(y, y_0)\}\mu(\mathrm{d}y) < \infty \quad \text{for some, and then for all, } y_0 \in \boldsymbol{Y}.$$

Let $\mathcal{P}_{\exp}(\boldsymbol{Y})$ be the family of such probability measures with, clearly, $\mathcal{P}_{\exp}(\boldsymbol{Y}) \subseteq \mathcal{P}_1(\boldsymbol{Y}) \subseteq \mathcal{P}(\boldsymbol{Y})$.

The Wasserstein distance between μ and ν in $\mathcal{P}_1(\boldsymbol{Y})$ (also referred to as the Kantorovich–Rubinstein metric) is defined as

$$\mathcal{W}(\mu, \nu) = \sup\left\{\int_{\boldsymbol{Y}} f \,\mathrm{d}\mu - \int_{\boldsymbol{Y}} f \,\mathrm{d}\nu\right\},$$

where the supremum ranges over all functions $f\colon \boldsymbol{Y} \to \mathbb{R}$ which are 1-Lipschitz continuous (note that the above integrals are finite due to $\mu, \nu \in \mathcal{P}_1(\boldsymbol{Y})$).

2. The control model $\mathcal{M}$: definition and assumptions

The main goal of this section is to introduce the notation, the parameters defining the model, and to present the construction of the controlled process. In particular, we construct a canonical measurable space $(\Omega, \mathcal{F})$ which describes the sample paths of the dynamic system. It is important to mention that, throughout this paper, we will deal with a controlled Markov process with *bounded cost and transition rates*; see the definition of $\mathcal{M}$ below.

2.1. Elements of the control model $\mathcal{M}$

We deal with a control model $\mathcal{M} = \{\boldsymbol{X}, \boldsymbol{A}, \{\boldsymbol{A}(x)\}_{x \in \boldsymbol{X}}, q, c\}$ with the following elements.

- The state space $\boldsymbol{X}$ is a Borel space with metric $d_{\boldsymbol{X}}$.
- The action space $\boldsymbol{A}$ is a Borel space with metric $d_{\boldsymbol{A}}$. The set of feasible actions in state $x \in \boldsymbol{X}$ is $\boldsymbol{A}(x)$, which is a nonempty measurable subset of $\boldsymbol{A}$. The set of admissible state-action pairs is

  $$\boldsymbol{K} = \{(x, a) \in \boldsymbol{X} \times \boldsymbol{A}\colon a \in \boldsymbol{A}(x)\} \in \mathcal{B}(\boldsymbol{X} \times \boldsymbol{A}).$$

  It is assumed that $\boldsymbol{K}$ contains the graph of a measurable function from $\boldsymbol{X}$ to $\boldsymbol{A}$. The multifunction from $\boldsymbol{X}$ to $\boldsymbol{A}$ given by $x \mapsto \boldsymbol{A}(x)$ will be denoted by Ψ.
- The transition rate q is a signed kernel on $\boldsymbol{X}$ given $\boldsymbol{K}$. This means that $B \mapsto q(B \mid x, a)$ is a signed measure on $(\boldsymbol{X}, \mathcal{B}(\boldsymbol{X}))$ for all $(x, a) \in \boldsymbol{K}$, and that $(x, a) \mapsto q(B \mid x, a)$ is

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

measurable for all $B \in \mathcal{B}(\boldsymbol{X})$. It satisfies $q(B \mid x, a) \geq 0$ for all $B \in \mathcal{B}(\boldsymbol{X})$ such that $x \notin B$. We assume that the transition kernel q is conservative, that is,

$$q(\boldsymbol{X} \mid x, a) = 0 \quad \text{for any } (x, a) \in \boldsymbol{K},$$

and that the transition rates are bounded: that is, for some constant $\hat{q}$

$$\sup_{(x,a)\in\boldsymbol{K}} \{-q(\{x\} \mid x, a)\} = \hat{q} > 0. \tag{2}$$

- The cost rate function is bounded and measurable: $c \in \mathbb{B}(\boldsymbol{K})$.

2.2. Construction of the process

Let $\boldsymbol{X}_\infty = \boldsymbol{X} \cup \{x_\infty\}$, where x_∞ is an isolated point. We let

$$\Omega_n = \boldsymbol{X} \times (\mathbb{R}_+^* \times \boldsymbol{X})^n \times (\{\infty\} \times \{x_\infty\})^\infty.$$

The canonical space, denoted by Ω, is defined as

$$\Omega = (\boldsymbol{X} \times (\mathbb{R}_+^* \times \boldsymbol{X})^\infty) \cup \bigcup_{n=1}^{\infty} \Omega_n$$

and it is endowed with its Borel σ-algebra, denoted by $\mathcal{F}$. For notational convenience, $\omega \in \Omega$ will be written as

$$\omega = (x_0, \theta_1, x_1, \theta_2, x_2, \ldots).$$

The canonical space Ω is given the following interpretation. Let $x_0 \in \boldsymbol{X}$ be the initial state of the dynamic system. Given $n \geq 0$, if $x_n \in \boldsymbol{X}$ then

- either $0 < \theta_{n+1} < \infty$, and we interpret θ_{n+1} as the sojourn time in state $x_n \in \boldsymbol{X}$, while $x_{n+1} \in \boldsymbol{X}$ is the post-jump location of the process;
- or $\theta_{n+1} = \infty$; this means that the dynamic system has been absorbed by x_n. In this case, we let $x_m = x_\infty$ and $\theta_m = \infty$ for all $m > n$. Such sample paths belong to Ω_n.

For every $n \in \mathbb{N}$ and $\omega \in \Omega$, let

$$h_n = (x_0, \theta_1, x_1, \theta_2, x_2, \ldots, \theta_n, x_n)$$

be the path up to n (we do not make ω explicit in the notation), and denote the collection of all such paths by $\boldsymbol{H}_n$.

For $n \in \mathbb{N}$, define the mapping $X_n : \Omega \to \boldsymbol{X}_\infty$ as $X_n(\omega) = x_n$. For $n \geq 1$, define Θ_n and T_n from Ω to $\overline{\mathbb{R}}_+^*$ as

$$\Theta_n(\omega) = \theta_n \quad \text{and} \quad T_n(\omega) = \theta_1 + \cdots + \theta_n.$$

We make the convention that $\Theta_0(\omega) = T_0(\omega) = 0$ for all $\omega \in \Omega$. Define $T_\infty(\omega) = \lim_{n\to\infty} T_n(\omega)$. The random variable T_∞ is referred to as the explosion time of the process. We denote by $H_n = (X_0, \Theta_1, X_1, \ldots, \Theta_n, X_n)$ the n-term history process, which takes values in $\boldsymbol{H}_n$ for $n \in \mathbb{N}$.

The random measure ν associated with $(\Theta_n, X_n)_{n\in\mathbb{N}}$ is a measure defined on $\mathbb{R}_+^* \times \boldsymbol{X}$ by

$$\nu(\omega;\ \mathrm{d}t, \mathrm{d}x) = \sum_{n\geq 1} \mathbf{1}_{\{T_n(\omega)<\infty\}}\, \delta_{(T_n(\omega),\, X_n(\omega))}(\mathrm{d}t, \mathrm{d}x).$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Informally, we can say that $\nu(\omega;\ \mathrm{d}t, \mathrm{d}x)$ puts a mass equal to 1 in each pair $(\theta_1 + \cdots + \theta_n, x_n)$ provided that all $\theta_1, \ldots, \theta_n$ are finite. For notational convenience, the dependence on ω will be suppressed and we will simply write $\nu(\mathrm{d}t, \mathrm{d}x)$. Define

$$\mathcal{F}_t = \sigma\{H_0\} \vee \sigma\{\nu((0,s] \times B)\colon s \le t,\ B \in \mathcal{B}(\boldsymbol{X})\} \quad \text{for } t \ge 0.$$

Finally, the continuous-time process $\{\xi_t\}_{t\ge 0}$ with values in $\boldsymbol{X}_\infty$ is given by

$$\xi_t(\omega) = \begin{cases} X_n(\omega) & \text{if } T_n(\omega) \le t < T_{n+1}(\omega) \text{ for } n \in \mathbb{N}, \\ x_\infty & \text{if } t \ge T_\infty(\omega). \end{cases}$$

So, the process $\{\xi_t\}_{t\ge 0}$ can be equivalently described by the sequence $(\Theta_n, X_n)_{n\in\mathbb{N}}$.

2.3. Admissible policies and distribution of the controlled process

Define $\boldsymbol{A}_\infty = \boldsymbol{A} \cup \{a_\infty\}$, where a_∞ is an isolated action associated to the cemetery state x_∞, and let $\boldsymbol{A}(x_\infty) = \{a_\infty\}$. We can extend the transition rate q to be a signed kernel on $\boldsymbol{X}_\infty$, given $\boldsymbol{K} \cup \{(x_\infty, a_\infty)\}$, by letting $q(\{x_\infty\} \mid x, a) = 0$ for all $(x,a) \in \boldsymbol{K}$ and $q(\cdot \mid x_\infty, a_\infty) \equiv 0$. We define q^+ as in (8).

An admissible control policy is a sequence $u = (\pi_n)_{n\in\mathbb{N}}$ where, for any $n \in \mathbb{N}$, π_n is a stochastic kernel (or transition probability measure) on $\boldsymbol{A}_\infty$, given $\boldsymbol{H}_n \times \mathbb{R}_+^*$, satisfying

$$\pi_n(\boldsymbol{A}(x_n) \mid h_n, t) = 1 \quad \text{for any } h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \boldsymbol{H}_n,\ t \in \mathbb{R}_+^*.$$

The set of admissible control policies is denoted by $\mathcal{U}$.

Given an admissible control policy $u = (\pi_n)_{n\in\mathbb{N}}$, we denote by π the random process with values in $\mathcal{P}(\boldsymbol{A}_\infty)$ as

$$\pi(\mathrm{d}a \mid t) = \sum_{n\in\mathbb{N}} \mathbf{1}_{\{T_n < t \le T_{n+1}\}}\, \pi_n(\mathrm{d}a \mid H_n, t - T_n) + \mathbf{1}_{\{t \ge T_\infty\}}\, \delta_{a_\infty}(\mathrm{d}a) \quad \text{for } t > 0. \tag{3}$$

It follows that π is an $\{\mathcal{F}_t\}_{t\in\mathbb{R}_+}$-predictable random process with values in $\mathcal{P}(\boldsymbol{A}_\infty)$.

Suppose that a control policy $u = (\pi_n)_{n\in\mathbb{N}} \in \mathcal{U}$ is fixed. We introduce the intensity of the jumps

$$\lambda_n(\Gamma, h_n, t) = \int_{\boldsymbol{A}_\infty} q^+(\Gamma \mid x_n, a)\pi_n(\mathrm{d}a \mid h_n, t),$$

and the rate of the jumps

$$\Lambda_n(\Gamma, h_n, t) = \int_0^t \lambda_n(\Gamma, h_n, s)\, \mathrm{d}s$$

for any $n \in \mathbb{N}$, $\Gamma \in \mathcal{B}(\boldsymbol{X}_\infty)$, $h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \boldsymbol{H}_n$, and $t \in \overline{\mathbb{R}}_+^*$. Now, for any $n \in \mathbb{N}$, the stochastic kernel G_n on $\overline{\mathbb{R}}_+^* \times \boldsymbol{X}_\infty$, given $\boldsymbol{H}_n$, is defined by

$$\begin{aligned} G_n(\Gamma \mid h_n) = {} & \delta_{(\infty, x_\infty)}(\Gamma)[\delta_{x_n}(\{x_\infty\}) + \delta_{x_n}(\boldsymbol{X})\mathrm{e}^{-\Lambda_n(\boldsymbol{X}, h_n, \infty)}] \\ & + \delta_{x_n}(\boldsymbol{X}) \int_{\Gamma \cap (\mathbb{R}_+^* \times \boldsymbol{X})} \lambda_n(\mathrm{d}x, h_n, t)\mathrm{e}^{-\Lambda_n(\boldsymbol{X}, h_n, t)}\, \mathrm{d}t \end{aligned}$$

for any $\Gamma \in \mathcal{B}(\overline{\mathbb{R}}_+^* \times \boldsymbol{X}_\infty)$ and $h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \boldsymbol{H}_n$.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Consider an admissible policy $u \in \mathcal{U}$ and an initial state $x \in \boldsymbol{X}$. From [13, Remark 3.43], there exists a probability $\mathbb{P}^{x,u}$ on $(\Omega, \mathcal{F})$ such that

$$\mathbb{P}^{x,u}\{X_0 = x\} = 1$$

and such that, for $\Gamma \in \mathcal{B}(\overline{\mathbb{R}}_+^* \times \boldsymbol{X}_\infty)$ and $n \geq 0$,

$$\mathbb{P}^{x,u}\{(\Theta_{n+1}, X_{n+1}) \in \Gamma \mid H_n\} = G_n(\Gamma \mid H_n) \quad \text{almost surely.}$$

We denote by $\mathbb{E}^{x,u}$ the expectation operator associated to $\mathbb{P}^{x,u}$.

Remark 1. Under the hypothesis that the transition kernel q is bounded (recall (2)), the continuous-time process $\{\xi_t\}_{t\geq 0}$ is nonexplosive under $\mathbb{P}^{x,u}$, which means that

$$\mathbb{P}^{x,u}\{T_\infty = \infty\} = 1 \quad \text{for any } x \in \boldsymbol{X},\ u \in \mathcal{U};$$

see, e.g. [14, Theorem 1].

2.4. Optimality criterion

Now we introduce the infinite-horizon performance criteria we are concerned with. Let $0 < \alpha < 1$ be a given discount factor. The total expected α-discounted cost of an admissible control policy $u \in \mathcal{U}$ for the initial state $x \in \boldsymbol{X}$ is defined as

$$\mathcal{V}_\alpha(u, x) = \mathbb{E}^{x,u}\left[\int_0^\infty \mathrm{e}^{-\alpha s} \int_{\boldsymbol{A}(\xi_s)} c(\xi_s, a)\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\right] \tag{4}$$

with π as in (3), and the long-run expected average cost of the control policy $u \in \mathcal{U}$ for the initial state $x \in \boldsymbol{X}$ is given by

$$\mathcal{J}(u, x) := \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{x,u}\left[\int_0^t \int_{\boldsymbol{A}(\xi_s)} c(\xi_s, a)\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\right]. \tag{5}$$

Our assumptions below will ensure that (4) and (5) are well defined and finite. The value function of the α-discounted control problem is

$$\mathcal{V}_\alpha^*(x) := \inf_{u\in\mathcal{U}} \mathcal{V}_\alpha(x, u) \quad \text{for } x \in \boldsymbol{X}.$$

Similarly, the value function of the average cost control problem is

$$\mathcal{J}^*(x) := \inf_{u\in\mathcal{U}} \mathcal{J}(x, u) \quad \text{for } x \in \boldsymbol{X},$$

and a policy $u^* \in \mathcal{U}$ is average cost optimal if $\mathcal{J}(x, u^*) = \mathcal{J}^*(x)$ for every $x \in \boldsymbol{X}$.

A control policy $u \in \mathcal{U}$ is called *deterministic stationary* if $\pi_n(\cdot \mid h_n, t) = \delta_{f(x_n)}(\cdot)$, where $f : \boldsymbol{X}_\infty \to \boldsymbol{A}_\infty$ is a measurable mapping satisfying $f(y) \in \boldsymbol{A}(y)$ for any $y \in \boldsymbol{X}_\infty$. We denote by F be the family of such measurable functions. By hypothesis, F is nonempty.

2.5. Assumptions and basic results

In this section we state our main assumptions on the control model $\mathcal{M}$, namely, Assumptions 1 and 2 below. These assumptions include the usual Lyapunov conditions and continuity-compactness requirements (which can be found in various forms in, e.g. [9], [10], [14], and [21]) plus some additional conditions ensuring that the solutions of the optimality equations are smooth enough.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Assumption 1 we adopt the following notation. Let Q be the stochastic kernel on $\boldsymbol{X}$, given $\boldsymbol{K}$, defined as

$$Q(\mathrm{d}y \mid x, a) = \frac{1}{\hat{q}} q(\mathrm{d}y \mid x, a) + \delta_x(\mathrm{d}y) \quad \text{for any } (x, a) \in \boldsymbol{K}.$$

Note that (2) implies that $Q(B \mid x, a) \geq 0$ for any $B \in \mathcal{B}(\boldsymbol{X})$ and $(x, a) \in \boldsymbol{K}$. In addition, since q is conservative, we indeed have $Q(\boldsymbol{X} \mid x, a) = 1$ for any $(x, a) \in \boldsymbol{K}$.

Assumption 1. (i) *The cost function c is L_c-Lipschitz continuous on $\boldsymbol{K}$.*

(ii) *The multifunction $\Psi : x \to \boldsymbol{A}(x)$ is compact valued and L_Ψ-Lipschitz continuous with respect to the Hausdorff distance.*

(iii) *For any $v \in \mathbb{C}(\boldsymbol{X})$, we have $qv \in \mathbb{C}(\boldsymbol{K})$.*

(iv) *There exists $L_Q > 0$ satisfying $(L_\Psi + 1)L_Q < 1$ such that, for every L_v-Lipschitz continuous function $v \in \mathbb{L}(\boldsymbol{X})$, we have $Qv \in \mathbb{L}(\boldsymbol{K})$ with Lipschitz constant $L_{Qv} = L_Q L_v$.*

The property in (iii) is usually referred to as the kernel q being weakly continuous, whereas in (iv) is referred to as the stochastic kernel Q being L_Q-Lipschitz continuous; see [12].

Assumption 2. *There is a function $w : \boldsymbol{X} \to [1, \infty)$ with the following properties:*

(i) *$w \in \mathbb{L}(\boldsymbol{X})$ and there exist constants $\rho > 0$ and $\gamma \geq 0$ with $qw(x, a) \leq -\rho w(x) + \gamma$ for any $(x, a) \in \boldsymbol{K}$;*

(ii) *there exists $x_0 \in \boldsymbol{X}$ such that the relative difference of the optimal discounted value function $h_\alpha(x) := \mathcal{V}^*_\alpha(x) - \mathcal{V}^*_\alpha(x_0)$ satisfies*

$$\sup_{\alpha > 0} \|h_\alpha\|_w < \infty.$$

Note that Assumptions 1(iv) and 2(i) imply that qw is continuous on $\boldsymbol{K}$.

Remark 2. The requirement in Assumption 2(ii) is a standard technical condition when dealing with average cost Markov controlled processes. Details can be found in [9, Assumption C], [10, Assumption C(ii)], and also in a slightly different form in [21, Condition 2]. Several sufficient (more easily verifiable, based on the primitive data of the control model) conditions for Assumption 2(ii) have been proposed in the literature. We refer the reader to [9, Lemma 3.3] where the authors proposed sufficient conditions based on uniform ergodicity properties, and also on drift and monotonicity conditions. Also, Guo and Ye [10, Theorem 3.3] proposed weak sufficient conditions based on communication properties and hitting times of the process. Finally, the discussion in [21, p. 959] yields a sufficient condition for Assumption 2(ii) based on the existence of a solution to a suitably defined drift inequality which is, in fact, closely related to the approach of Guo and Ye [10, Theorem 3.3].

Before stating our main result on the average cost optimality inequalities, we prove a preliminary fact.

Lemma 1. *Under Assumptions 1 and 2, there exists a constant $L_{\mathcal{V}^*}$ such that*

$$|\mathcal{V}^*_\alpha(x) - \mathcal{V}^*_\alpha(y)| \leq L_{\mathcal{V}^*} d_{\boldsymbol{X}}(x, y) \quad \text{for any } x, y \in \boldsymbol{X},\ \alpha > 0.$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Proof. Standard arguments (see, e.g. [14, Theorem 4]) can be used to show that the value function of the α-discounted control problem $\mathcal{V}_\alpha^*(x)$ is equal to $\lim_{k\to\infty} \boldsymbol{W}_{\alpha,k}(x)$ for any $\alpha > 0$ and $x \in \boldsymbol{X}$, where

$$\boldsymbol{W}_{\alpha,k+1}(x) = \inf_{a\in\boldsymbol{A}(x)} \left\{ \frac{c(x,a)}{\hat{q}+\alpha} + \frac{\hat{q}}{\hat{q}+\alpha} Q\boldsymbol{W}_{\alpha,k}(x,a) \right\},$$

and $\boldsymbol{W}_{\alpha,0}(x) = 0$. It can be easily shown by induction that, for any $\alpha > 0$ and $k \in \mathbb{N}$, the mappings $\boldsymbol{W}_{\alpha,k}$ are $L_{\boldsymbol{W}_{\alpha,k}}$-Lipschitz continuous with

$$L_{\boldsymbol{W}_{\alpha,k+1}} = \left(\frac{L_c}{\hat{q}} + L_Q L_{\boldsymbol{W}_{\alpha,k}} \right)(1 + L_\Psi).$$

Combining the previous equation and Assumption 1(iv), it follows that

$$L_{\boldsymbol{W}_{\alpha,k}} \le \frac{L_c}{\overline{q}} \frac{1+L_\Psi}{1-L_Q(1+L_\Psi)} \quad \text{for any } \alpha > 0,\ k \in \mathbb{N}.$$

Now recalling that $\mathcal{V}_\alpha^*(x) = \lim_{k\to\infty} \boldsymbol{W}_{\alpha,k}(x)$, we obtain the result with

$$L_{\mathcal{V}^*} = \frac{L_c}{\overline{q}} \frac{1+L_\Psi}{1-L_Q(1+L_\Psi)}$$

the Lipschitz constant of $\mathcal{V}_\alpha^*$. □

The following result is essentially the same as [9, Theorem 4.2]. It differs by the fact that, on the one hand, we consider a weakly continuous bounded transition kernel q that does not satisfy the strong continuity property imposed by [9, Assumption B2]; and, on the other hand, we show that the functions v_1 and v_2 involved in the optimality inequalities are Lipschitz continuous by imposing stronger continuity properties on parameters of the model $\mathcal{M}$. We provide only a sketch of the proof, which mainly combines Lemma 1 and arguments from [9, Theorem 4.2] and [21, Theorem 1].

Theorem 1. *Suppose that the control model $\mathcal{M}$ satisfies Assumptions 1 and 2.*

(i) *There exist a constant g^* and functions $v_1, v_2 \in \mathbb{L}_w(\boldsymbol{X})$ that are solutions of the average optimality inequalities:*

$$g^* \ge \inf_{a\in\boldsymbol{A}(x)} \left\{ c(x,a) + \int_{\boldsymbol{X}} v_1(y) q(\mathrm{d}y \mid x,a) \right\}, \tag{6}$$

$$g^* \le \inf_{a\in\boldsymbol{A}(x)} \left\{ c(x,a) + \int_{\boldsymbol{X}} v_2(y) q(\mathrm{d}y \mid x,a) \right\} \tag{7}$$

for every $x \in \boldsymbol{X}$.

(ii) *The optimal average cost of $\mathcal{M}$ is constant and $g^* = \mathcal{J}^*(x)$ for all $x \in \boldsymbol{X}$.*

(iii) *Any $f \in F$ attaining the infimum in (6) is average optimal for $\mathcal{M}$, and such an f indeed exists.*

Proof. (i) Under our standing assumptions, it can be shown (see, e.g. [14, Theorem 4]) that the function h_α introduced in Assumption 2(ii) satisfies the following equation:

$$\frac{\alpha \mathcal{V}_\alpha^*(x_0)}{\hat{q}} + \frac{\alpha h_\alpha(x)}{\hat{q}} + h_\alpha(x) = \inf_{a\in\boldsymbol{A}(x)} \left\{ \frac{c(x,a)}{\hat{q}} + Qh_\alpha(x,a) \right\}.$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Define

$$g^* = \limsup_{\alpha \to 0} \alpha \mathcal{V}_\alpha^*(x_0) \quad \text{and} \quad v_1(x) = \liminf_{(\alpha, y) \to (0, x)} h_\alpha(y) \quad \text{for any } x \in \boldsymbol{X}.$$

Using similar arguments to those of [21, Theorem 1], we see that (6) is satisfied. Clearly, Assumption 2(ii) implies that $v_1 \in \mathbb{B}_w(\boldsymbol{X})$. The only point is to show that $v_1 \in \mathbb{L}_w(\boldsymbol{X})$. However, recalling from Lemma 1 the fact that

$$|\mathcal{V}_\alpha^*(x) - \mathcal{V}_\alpha^*(y)| \leq L_{\mathcal{V}^*} d_{\boldsymbol{X}}(x, y)$$

for any $x, y \in \boldsymbol{X}$ and any $\alpha > 0$, for a constant $L_{\mathcal{V}^*}$ not depending on α, it follows easily that $v_1 \in \mathbb{L}_w(\boldsymbol{X})$. The reasoning to show the existence of a function $v_2 \in \mathbb{B}_w(\boldsymbol{X})$ satisfying (7) is basically the same as for the proof of Equation (4.2) of [9, Theorem 4.2], since it does not use any continuity properties of the transition kernel. Again, an application of Lemma 1 implies that $v_2 \in \mathbb{L}_w(\boldsymbol{X})$.

The proof of (ii) and (iii) is identical to statements (b) and (c) of [9, Theorem 4.2]. □

3. Approximation of the optimal value of $\mathcal{M}$

The objective of this section is to address the issue of the approximation of the optimal value of control model $\mathcal{M}$. It is based on a suitable discretization of the state and action spaces. To achieve this discretization, we require additional conditions on $\mathcal{M}$, which are contained in Assumption 3.

Denote by q^+ the positive part of the transition kernel q; that is, consider the transition measure on $\boldsymbol{X}$, given $\boldsymbol{K}$, defined for $B \in \mathcal{B}(\boldsymbol{X})$ and $(x, a) \in \boldsymbol{K}$ as

$$q^+(B \mid x, a) = q(B - \{x\} \mid x, a).$$

We can write, equivalently,

$$q(\mathrm{d}y \mid x, a) = q^+(\mathrm{d}y \mid x, a) + q(\{x\} \mid x, a)\delta_x(\mathrm{d}y). \tag{8}$$

Next we impose the condition that the kernel q^+ is absolutely continuous with respect to some probability measure μ for a sufficiently regular density function. Assumption 3(i)–(iii) will be useful for the discretization of the state space, while Assumption 3(iv) will be the basis for the discretization of the action sets.

Assumption 3. *There exist a probability measure $\mu \in \mathcal{P}_1(\boldsymbol{X})$ and a nonnegative function p defined on $\boldsymbol{X} \times \boldsymbol{K}$ such that the following hold:*

(i) *for all $B \in \mathcal{B}(\boldsymbol{X})$ and $(x, a) \in \boldsymbol{K}$, we have*

$$q^+(B \mid x, a) = \int_B p(y \mid x, a)\mu(\mathrm{d}y);$$

(ii) *there exists some $L_p > 0$ such that the function $p(\cdot \mid x, \cdot)$ is L_p-Lipschitz continuous on $\boldsymbol{X} \times \boldsymbol{A}(x)$ for any $x \in \boldsymbol{X}$;*

(iii) *for some positive constants $\hat{p}$ and L_{wp}, we have*

$$p(y \mid x, a) \leq \hat{p} w(x) \quad \text{and} \quad |w(y)p(y \mid x, a) - w(z)p(z \mid x, a)| \leq L_{wp} w(x) d_{\boldsymbol{X}}(y, z)$$

for $y, z \in \boldsymbol{X}$ and $(x, a) \in \boldsymbol{K}$, where the function w comes from Assumption 2;

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

(iv) *for each* $\delta > 0$ *and every* $x \in \boldsymbol{X}$, *there is a finite set* $\boldsymbol{A}_\delta(x) \subseteq \boldsymbol{A}(x)$ *such that the multifunction defined on* $\boldsymbol{X}$ *by* $x \mapsto \boldsymbol{A}_\delta(x)$ *is Borel-measurable and*

$$d_H(\boldsymbol{A}(x), \boldsymbol{A}_\delta(x)) \leq \delta w(x).$$

Before addressing the issue of the approximation of the control model $\mathcal{M}$, we state a technical result that will be useful in what follows. It deals with a Lipschitz continuity property of the density function p introduced in Assumption 3.

Lemma 2. *Suppose that Assumptions 2 and 3 hold. Given* $v \in \mathbb{L}_w(X)$ *and* $(x, a) \in \boldsymbol{K}$, *the function* $y \mapsto v(y)p(y \mid x, a)$ *is Lipschitz continuous on* $\boldsymbol{X}$ *with Lipschitz constant given by* $\mathcal{L}_{vp}w(x)$ *with*

$$\mathcal{L}_{vp} := \|v\|_w(L_{wp} + \hat{p}L_w) + L_v\hat{p}.$$

Proof. See [7, Lemma 3.3]. □

3.1. Construction of the approximating model $\mathcal{M}_{k,\delta}$

In order to approximate the control model $\mathcal{M}$, we will introduce a new control model $\mathcal{M}_{k,\delta}$ depending on the parameters of discretizations k and δ. Loosely speaking, the technique to construct $\mathcal{M}_{k,\delta}$ will consist of the following steps.

- Discretizing the state space $\boldsymbol{X}$ by *replacing* the measure μ in Assumption 3(i) by a probability measure μ_k supported on k points in $\boldsymbol{X}$. The accuracy of the approximation of μ_k to μ will be measured in terms of the distance $\mathcal{W}(\mu, \mu_k)$, which will be assumed to converge to 0 as $k \to \infty$; see the discussions in Sections 4.1 and 4.2.
- Discretizing the action sets $\boldsymbol{A}(x)$ by considering the finite δ-nets $\boldsymbol{A}_\delta(x)$ in Assumption 3(iv). The approximation of the action sets will become more accurate as $\delta \to 0$.

We now introduce the model $\mathcal{M}_{k,\delta}$ and its associated parameters.

Definition 1. For $k \geq 1$ and $\delta > 0$, let μ_k be a probability measure on $\boldsymbol{X}$ with finite support and let $\boldsymbol{A}_\delta(x)$ a finite subset of $\boldsymbol{A}(x)$ for any $x \in \boldsymbol{X}$. The elements of the control model $\mathcal{M}_{k,\delta}$ are given by $\{\boldsymbol{X}, \boldsymbol{A}, \{\boldsymbol{A}_\delta(x)\}_{x\in\boldsymbol{X}}, q_k, c\}$ with q_k defined as

$$q_k(B \mid x, a) = \int_B p(y \mid x, a)\mu_k(\mathrm{d}y) - \delta_x(B)\int_{\boldsymbol{X}} p(y \mid x, a)\mu_k(\mathrm{d}y)$$

for $B \in \mathcal{B}(\boldsymbol{X})$ and $(x, a) \in \boldsymbol{K}_\delta$, where

$$\boldsymbol{K}_\delta = \{(x, a) \in \boldsymbol{X} \times \boldsymbol{A} \colon a \in \boldsymbol{A}_\delta(x)\} \in \mathcal{B}(\boldsymbol{K}).$$

We can extend the transition rate q_k to be a signed kernel on $\boldsymbol{X}_\infty$, given $\boldsymbol{K}_\delta \cup \{(x_\infty, a_\infty)\}$, by letting $q_k(\{x_\infty\} \mid x, a) = 0$ for all $(x, a) \in \boldsymbol{K}_\delta$ and $q_k(\cdot \mid x_\infty, a_\infty) \equiv 0$. An admissible control policy for the model $\mathcal{M}_{k,\delta}$ is a sequence $u = (\pi_n)_{n\in\mathbb{N}}$, where for any $n \in \mathbb{N}$, π_n is a stochastic kernel (or transition probability measure) on $\boldsymbol{A}_\infty$, given $\boldsymbol{H}_n \times \mathbb{R}_+^*$, satisfying

$$\pi_n(\boldsymbol{A}_\delta(x_n) \mid h_n, t) = 1 \quad \text{for any } h_n = (x_0, \theta_1, x_1, \ldots, \theta_n, x_n) \in \boldsymbol{H}_n,\ t \in \mathbb{R}_+^*.$$

The set of admissible control policies for the model $\mathcal{M}_{k,\delta}$ is denoted by $\mathcal{U}_\delta$. Observe that $\mathcal{U}_\delta \subseteq \mathcal{U}$. Consider an admissible policy $u = (\pi_n)_{n\in\mathbb{N}} \in \mathcal{U}_\delta$ and an initial state $x \in \boldsymbol{X}$. Similarly to the construction described in Sections 2.2 and 2.3, there exists a probability measure $\mathbb{P}_{k,\delta}^{x,u}$ on $(\Omega, \mathcal{F})$ that models $\mathcal{M}_{k,\delta}$. We write $\mathbb{E}_{k,\delta}^{x,u}$ for the expectation operator associated to $\mathbb{P}_{k,\delta}^{x,u}$ and denote by F_δ the family of measurable functions $\varphi \colon \boldsymbol{X}_\infty \to \boldsymbol{A}_\infty$ satisfying $\varphi(y) \in \boldsymbol{A}_\delta(y)$ for any $y \in \boldsymbol{X}_\infty$.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Note that $\mathcal{M}_{k,\delta}$ is essentially a finite state control model because, starting from any initial state $x \in \boldsymbol{X}$, after the first transition the state of the system belongs to (and remains in) the support of μ_k. The long-run expected average cost of the control policy $u \in \mathcal{U}_\delta$ for the initial state $x \in \boldsymbol{X}$ is given by

$$\mathcal{J}_{k,\delta}(x,u) := \limsup_{t\to\infty} \frac{1}{t}\mathbb{E}^{x,u}_{k,\delta}\left[\int_0^t \int_{\boldsymbol{A}_\delta(\xi_s)} c(\xi_s,a)\pi(\mathrm{d}a \mid s)\,\mathrm{d}s\right]$$

and the value function of the average cost control problem is

$$\mathcal{J}^*_{k,\delta}(x) := \inf_{u\in\mathcal{U}_\delta} \mathcal{J}_{k,\delta}(x,u) \quad \text{for } x \in \boldsymbol{X}.$$

For our next remark, we define q_k^+ as the positive part of the signed kernel q_k; see (8).

Remark 3. Observe that, for any $(x,a) \in \boldsymbol{K}_\delta$,

$$\begin{aligned} q_k^+(\boldsymbol{X} \mid x,a) &= \int_{\boldsymbol{X}} p(y \mid x,a)\mu_k(\mathrm{d}y) \\ &\le \int_X p(y \mid x,a)\mu(\mathrm{d}y) + L_p\mathcal{W}(\mu,\mu_k) \\ &= q^+(\boldsymbol{X} \mid x,a) + L_p\mathcal{W}(\mu,\mu_k) \\ &\le \hat{q} + L_p\mathcal{W}(\mu,\mu_k), \end{aligned}$$

where we make use of (2) and Assumption 3(ii). So, the continuous-time process $\{\xi_t\}_{t\ge 0}$ has bounded transition rates and it is, therefore, nonexplosive under $\mathbb{P}^{x,u}_{k,\delta}$ for any $x \in \boldsymbol{X}$ and $u \in \mathcal{U}_\delta$; see, e.g. [14, Theorem 1]. Hence, $\mathcal{J}^*_{k,\delta}(x)$ is well defined and finite.

3.2. Estimation of the value function

Having introduced the model $\mathcal{M}_{k,\delta}$, we now study the problem of approximating the value function $\mathcal{J}^*(x)$ for the original control model $\mathcal{M}$.

In our main result in this section we establish that the difference $|g^* - \mathcal{J}^*_{k,\delta}(x)|$ between the optimal average costs of $\mathcal{M}$ and $\mathcal{M}_{k,\delta}$ is bounded by a linear combination of the Wasserstein distance $\mathcal{W}(\mu,\mu_k)$ and the parameter δ in Assumption 3(iv). Our results in this section are presented in a general framework and, for the moment, we do not impose any particular conditions on the measures $\{\mu_k\}_{k\in\mathbb{N}}$ and the sets $\{\boldsymbol{A}_\delta(x)\}_{x\in\boldsymbol{X}}$, other than those in Definition 1.

In our next result we compare q with q_k, when applied to a Lipschitz continuous function, and show that the model $\mathcal{M}_{k,\delta}$ inherits the Lyapunov condition satisfied by the model $\mathcal{M}$ (see Assumption 2(i)) provided $\mathcal{W}(\mu,\mu_k)$ is *small*.

Lemma 3. *Suppose that $\mathcal{M}$ satisfies Assumptions 1–3, and let $k \ge 1$ and $\delta > 0$.*

(i) *Given $(x,a) \in \boldsymbol{K}_\delta$ and $v \in \mathbb{L}_w(\boldsymbol{X})$, we have*

$$\left|\int_{\boldsymbol{X}} v(y)q(\mathrm{d}y \mid x,a) - \int_{\boldsymbol{X}} v(y)q_k(\mathrm{d}y \mid x,a)\right| \le [\mathcal{L}_{vp} + L_p\|v\|_w]w(x)\mathcal{W}(\mu,\mu_k).$$

(ii) *For all $(x,a) \in \boldsymbol{K}_\delta$,*

$$\int_{\boldsymbol{X}} w(y)q_k(\mathrm{d}y \mid x,a) \le -[\rho - (L_{wp} + L_p)\mathcal{W}(\mu,\mu_k)]w(x) + \gamma.$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Proof. (i) Clearly, from Assumption 3(i),

$$\begin{aligned}\left|\int_{\boldsymbol{X}} v(y)q(\mathrm{d}y \mid x,a) - \int_{\boldsymbol{X}} v(y)q_k(\mathrm{d}y \mid x,a)\right| \\ \leq \left|\int_{\boldsymbol{X}} v(y)p(y \mid x,a)\mu(\mathrm{d}y) - \int_{\boldsymbol{X}} v(y)p(y \mid x,a)\mu_k(\mathrm{d}y)\right| \\ + |v(x)|\left|\int_{\boldsymbol{X}} p(y \mid x,a)\mu(\mathrm{d}y) - \int_{\boldsymbol{X}} p(y \mid x,a)\mu_k(\mathrm{d}y)\right|.\end{aligned}$$

The claim is then easy to obtain by using Lemma 2 and Assumptions 3(ii) and 3(iii).

(ii) Similarly, we have

$$\left|\int_{\boldsymbol{X}} w(y)q(\mathrm{d}y \mid x,a) - \int_{\boldsymbol{X}} w(y)q_k(\mathrm{d}y \mid x,a)\right| \leq [L_{wp} + L_p]w(x)\mathcal{W}(\mu,\mu_k),$$

from Assumptions 3(ii) and 3(iii). The result readily follows by using Assumption 2(i). □

In the following two results we see that if the control $\mathcal{M}$ model satisfies average optimality inequalities of the type (6) and (7) then the corresponding constant g^* can be approximated by $\mathcal{J}^*_{k,\delta}(x)$.

Proposition 1. *Suppose that $\mathcal{M}$ satisfies Assumptions 1–3. If there is a constant $g \in \mathbb{R}$ and a function $v \in \mathbb{L}_w(\boldsymbol{X})$ such that*

$$g \leq c(x,a) + \int_{\boldsymbol{X}} v(y)q(\mathrm{d}y \mid x,a) \quad \textit{for every } (x,a) \in \boldsymbol{K}, \tag{9}$$

then the average cost for the control model $\mathcal{M}_{k,\delta}$ satisfies

$$g \leq \mathcal{J}^*_{k,\delta}(x) + [\mathcal{L}_{vp} + L_p\|v\|_w]\mathcal{W}(\mu,\mu_k)\frac{2\gamma}{\rho} \quad \textit{for any } x \in \boldsymbol{X}$$

provided that $\mathcal{W}(\mu,\mu_k) \leq \rho/2(L_{wp+L_p})$.

Proof. Consider an admissible policy $u = (\pi_{\delta,n})_{n\in\mathbb{N}} \in \mathcal{U}_\delta$ with $\delta > 0$ and a function $v \in \mathbb{L}_w(\boldsymbol{X})$. From Remark 3, we see that the continuous-time process $\{\xi_t\}_{t\geq 0}$ is nonexplosive under $\mathbb{P}^{x,u}_{k,\delta}$, that is, $\mathbb{P}^{x,u}_{k,\delta}\{T_\infty = \infty\} = 1$ for any $x \in \boldsymbol{X}$ and $k \in \mathbb{N}$. For notational convenience, we denote by π_δ the random process with values in $\mathcal{P}(\boldsymbol{A}_\infty)$ as

$$\pi_\delta(\mathrm{d}a \mid t) = \sum_{n\in\mathbb{N}} \mathbf{1}_{\{T_n < t \leq T_{n+1}\}}\, \pi_{\delta,n}(\mathrm{d}a \mid H_n, t - T_n) \quad \text{for any } t > 0.$$

Now observe that

$$\int_{\boldsymbol{A}_\delta(\xi_s)} \left[c(\xi_s,a) + \int_{\boldsymbol{X}} v(y)q_k(\mathrm{d}y \mid \xi_s,a)\right]\pi_\delta(\mathrm{d}a \mid s)$$

is well defined. Consequently, from Lemma 3(i),

$$\begin{aligned}g \leq \int_{\boldsymbol{A}_\delta(\xi_s)} \left[c(\xi_s,a) + \int_{\boldsymbol{X}} v(y)q_k(\mathrm{d}y \mid \xi_s,a)\right]\pi_\delta(\mathrm{d}a \mid s) \\ + [\mathcal{L}_{vp} + L_p\|v\|_w]w(\xi_s)\mathcal{W}(\mu,\mu_k) \quad \text{for every } s \geq 0.\end{aligned}$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

From Remark 3 and Lemma 3(i), we can use Dynkin's formula (see, e.g. [14, Theorem 2]) to obtain

$$\begin{aligned}\mathbb{E}_{k,\delta}^{x,u}[v(\xi_t)] - v(x) &= \mathbb{E}_{k,\delta}^{x,u}\left[\int_0^t \int_{\boldsymbol{A}_\delta(\xi_s)} \int_{\boldsymbol{X}} v(y) q_k(\mathrm{d}y \mid \xi_s, a)\pi_\delta(\mathrm{d}a \mid s)\,\mathrm{d}s\right] \\ &\geq gt - \mathbb{E}_{k,\delta}^{x,u}\left[\int_0^t \int_{\boldsymbol{A}_\delta(\xi_s)} c(\xi_s, a)\pi_\delta(\mathrm{d}a \mid s)\,\mathrm{d}s\right] \\ &\quad - [\mathcal{L}_{vp} + L_p\|v\|_w]\mathcal{W}(\mu, \mu_k)\mathbb{E}_{k,\delta}^{x,u}\left[\int_0^t w(\xi_s)\,\mathrm{d}s\right]. \end{aligned} \tag{10}$$

By Lemma 3(ii) and using [14, Theorem 1(a)], we have

$$\mathbb{E}_{k,\delta}^{x,u_\delta}[w(\xi_s)] \leq \mathrm{e}^{-\rho t/2} w(x) + \frac{2\gamma}{\rho}$$

since $\mathcal{W}(\mu, \mu_k) \leq \rho/2(L_{wp+L_p})$ and so, dividing by t and taking the lim sup as $t \to \infty$ in (10), we obtain

$$\mathcal{J}_{k,\delta}(x, u_\delta) \geq g - [\mathcal{L}_{vp} + L_p\|v\|_w]\mathcal{W}(\mu, \mu_k)\frac{2\gamma}{\rho},$$

yielding the result. □

Proposition 2. *Suppose that $\mathcal{M}$ satisfies Assumptions 1–3. If there is a constant $g \in \mathbb{R}$, a function $v \in \mathbb{L}_w(\boldsymbol{X})$, and $f^* \in F$ such that*

$$g \geq c(x, f^*(x)) + \int_{\boldsymbol{X}} v(y) q(\mathrm{d}y \mid x, f^*(x)) \quad \textit{for every } x \in \boldsymbol{X}, \tag{11}$$

then there exists $f_\delta^* \in F_\delta$ such that the corresponding average cost $\mathcal{J}_{k,\delta}(x, f_\delta^*)$ for the control model $\mathcal{M}_{k,\delta}$ satisfies

$$g \geq \mathcal{J}_{k,\delta}(x, f_\delta^*) - ([L_c + \|v\|_w\mu(w)L_p]\delta + [\mathcal{L}_{vp} + L_p\|v\|_w]\mathcal{W}(\mu, \mu_k))\frac{2\gamma}{\rho} \quad \textit{for any } x \in \boldsymbol{X}$$

provided that $\mathcal{W}(\mu, \mu_k) \leq \rho/2(L_{wp+L_p})$.

Proof. From Assumption 3(iv), we have $\min_{a \in \boldsymbol{A}_\delta(x)} d_{\boldsymbol{A}}(a, f^*(x)) = d_{\boldsymbol{A}}(f_\delta^*(x), f^*(x))$ for some $f_\delta^* \in F_\delta$ and $d_{\boldsymbol{A}}(f_\delta^*(x), f^*(x)) \leq \delta w(x)$. Therefore,

$$\begin{aligned} g &\geq c(x, f_\delta^*(x)) + \int_{\boldsymbol{X}} v(y) q(\mathrm{d}y \mid x, f_\delta^*(x)) - |c(x, f^*(x)) - c(x, f_\delta^*(x))| \\ &\quad - \left|\int_{\boldsymbol{X}} v(y)[p(y \mid x, f^*(x)) - p(y \mid x, f_\delta^*(x))]\mu(\mathrm{d}y)\right| \\ &\geq c(x, f_\delta^*(x)) + \int_{\boldsymbol{X}} v(y) q(\mathrm{d}y \mid x, f_\delta^*(x)) - [L_c + \|v\|_w\mu(w)L_p]\delta w(x), \end{aligned}$$

where in the last inequality we used Assumptions 1(i) and 3(ii). Consequently,

$$\begin{aligned} g &\geq c(\xi_s, f_\delta^*(\xi_s)) + \int_{\boldsymbol{X}} v(y) q_k(\mathrm{d}y \mid \xi_s, f_\delta^*(\xi_s)) \\ &\quad - ([L_c + \|v\|_w\mu(w)L_p]\delta + [\mathcal{L}_{vp} + L_p\|v\|_w]\mathcal{W}(\mu, \mu_k))w(\xi_s). \end{aligned}$$

Proceeding as in the proof of the previous proposition, we obtain the result. □

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

From these two propositions, we can now provide a bound on the approximation error of the value function.

Theorem 2. *Suppose that Assumptions 1–3 hold. For any $x \in \boldsymbol{X}$, $\delta > 0$, and $k \geq 1$ such that $\mathcal{W}(\mu, \mu_k) \leq \rho/2(L_{wp+L_p})$, we have*

$$\sup_{x \in \boldsymbol{X}} |g^* - \mathcal{J}^*_{k,\delta}(x)| \leq \mathcal{C}\delta + \mathcal{D}\mathcal{W}(\mu, \mu_k),$$

where

$$\mathcal{C} = [L_c + \|v_1\|_w \mu(w) L_p] \frac{2\gamma}{\rho}, \qquad \mathcal{D} = ([\mathcal{L}_{v_1 p} + L_p \|v_1\|_w] \vee [\mathcal{L}_{v_2 p} + L_p \|v_2\|_w]) \frac{2\gamma}{\rho}.$$

Proof. From Theorem 1, there exist a constant g^* and a function $v_2 \in \mathbb{L}_w(\boldsymbol{X})$ satisfying (9). Therefore, applying Proposition 1, we obtain

$$g^* \leq \mathcal{J}^*_{k,\delta}(x) + [\mathcal{L}_{v_2 p} + L_p \|v_2\|_w] \mathcal{W}(\mu, \mu_k) \frac{2\gamma}{\rho} \quad \text{for any } \delta > 0$$

provided that $\mathcal{W}(\mu, \mu_k) \leq \rho/2(L_{wp+L_p})$. Now from items (i) and (ii) of Theorem 1, there exist a function $v_1 \in \mathbb{L}_w(\boldsymbol{X})$ and $f^* \in F$ satisfying (11) for the constant g^*, and so applying Proposition 2, we have

$$g^* \geq \mathcal{J}^*_{k,\delta}(x) - ([L_c + \|v_1\|_w \mu(w) L_p]\delta + [\mathcal{L}_{v_1 p} + L_p \|v_1\|_w] \mathcal{W}(\mu, \mu_k)) \frac{2\gamma}{\rho} \quad \text{for any } \delta > 0$$

provided that $\mathcal{W}(\mu, \mu_k) \leq \rho/2(L_{wp+L_p})$. This establishes the result. □

Therefore, we can indeed bound the approximation error $|g^* - \mathcal{J}^*_{k,\delta}(x)|$ by a linear combination of δ (the parameter of the discretization of the action space) and $\mathcal{W}(\mu, \mu_k)$, the distance between μ and μ_k. Note that $\mathcal{L}_{v_1 p}$ and $\mathcal{L}_{v_2 p}$ can be explicitly computed since they depend on constants given in our assumptions such as, e.g. $\mathcal{L}_{\mathcal{V}^*}$ derived in Lemma 1.

Finally, we make a remark on our approach used here for the approximation of the continuous-time model $\mathcal{M}$. In both Lemma 1 and Theorem 1 we have used the so-called *uniformization technique* to transform $\mathcal{M}$ into an equivalent discrete-time model. Namely, as shown in Lemma 1, the corresponding discounted discrete-time model has

$$\frac{c(x,a)}{\hat{q} + \alpha}, \qquad \frac{\hat{q}}{\hat{q} + \alpha}, \qquad Q(\mathrm{d}y \mid x, a) = \frac{1}{\hat{q}} q(\mathrm{d}y \mid x, a) + \delta_x(\mathrm{d}y)$$

as its respective cost function, discount factor, and transition probabilities. In Theorem 1 we used the vanishing discount technique to deal with the average optimality of this discrete-time model: letting the discount rate $\alpha \downarrow 0$ makes the discount factor $\hat{q}/(\hat{q} + \alpha) \uparrow 1$.

We would like to stress that it is not possible, in general, to apply the techniques developed in [8] to approximate the 'uniformized' discrete-time model in order to obtain the corresponding approximations for $\mathcal{M}$. The reason is that a key hypothesis of [8] is the existence of a probability measure ν on $\boldsymbol{X}$ and a density function $u(y \mid x, a)$ such that

$$Q(\mathrm{d}y \mid x, a) = \frac{1}{\hat{q}} q(\mathrm{d}y \mid x, a) + \delta_x(\mathrm{d}y) = u(y \mid x, a) \nu(\mathrm{d}y). \tag{12}$$

This implies that each $x \in \boldsymbol{X}$ such that $-q(\{x\} \mid x, a) < \hat{q}$ for some $a \in \boldsymbol{A}(x)$ (recall (2)) is necessarily an atom of ν. As a consequence, (12) entails that for almost all $x \in \boldsymbol{X}$ (excluding, perhaps, the atoms of ν) and all $a \in \boldsymbol{A}(x)$, the instantaneous jump rate $-q(\{x\} \mid x, a)$ is *constant*.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

By following the discrete-time approach of [8] for the continuous-time model herein, we would need to restrict ourselves to a really poor class of continuous-time models: the distribution of the sojourn times should be the same for almost all states, regardless of the controller's action. This is a fairly unnatural hypothesis in continuous-time modeling. This is the reason why, although our proof techniques use a discrete-time model, we indeed need to follow a specific approach for the continuous-time model.

4. Empirical and deterministic approximations

In this section we discuss the approximation of the underlying probability measure μ in the metric $\mathcal{W}$. Actually, two main approaches exist: one uses empirical distribution and the other one a so-called deterministic approximation.

4.1. The empirical approach

Now we specialize the main result (see Theorem 2) to the case when the measure μ_k is given by the empirical probability measure drawn from the measure μ.

First we recall some general results on the convergence in the Wasserstein metric of empirical probability measures. Let $\boldsymbol{Y}$ be a Borel space and consider the probability space $(\boldsymbol{Y}^\infty, \mathcal{B}(\boldsymbol{Y})^\infty, \mathbb{P}_\mu)$ which we define next. The elements of the product space $\boldsymbol{Y}^\infty$ are written as $\zeta = (\zeta_1, \zeta_2, \ldots)$. It follows that $\mathcal{B}(\boldsymbol{Y})^\infty$ is the product σ-algebra of the $\mathcal{B}(\boldsymbol{Y})$, that is, the minimal σ-algebra containing the rectangles

$$R = \{\zeta \in \boldsymbol{Y}^\infty \colon \zeta_1 \in B_1, \ldots, \zeta_k \in B_k\} \quad \text{for all } k \geq 1,\ B_1, \ldots, B_k \in \mathcal{B}(\boldsymbol{Y}).$$

The probability measure $\mathbb{P}_\mu$ is such that, for $R \in \mathcal{B}(\boldsymbol{Y})^\infty$ as above, $\mathbb{P}_\mu(R) = \mu(B_1) \cdots \mu(B_k)$. Given $n \geq 1$ and $\zeta \in \boldsymbol{Y}^\infty$, the empirical probability measure $\mu_n(\zeta)$ is the probability measure on $\boldsymbol{Y}$ defined as

$$\mu_n(\zeta) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\zeta_i}.$$

We will also say that μ_n is a random probability measure. In what follows we will use the following convergence result which yields the rate of convergence (in probability) of μ_n to μ in the Wasserstein metric; it is taken from [3, Corollary 2.5].

Proposition 3. *Let $\boldsymbol{Y}$ be a Borel space. Given $\mu \in \mathcal{P}_{\exp}(\boldsymbol{Y})$ and $\varepsilon > 0$ there exist positive constants C_ε and D_ε such that*

$$\mathbb{P}_\mu\{\mathcal{W}(\mu, \mu_n(\zeta)) > \varepsilon\} \leq C_\varepsilon \exp\{-D_\varepsilon n\} \quad \textit{for all } n \geq 1.$$

With the empirical approach described above, we know that, for a given precision $\varepsilon > 0$ and for large enough n, there is a small probability (decaying exponentially in n) that we are not being ε-precise when approximating μ with the random measure μ_n, which is supported on at most n points in $\boldsymbol{Y}$.

Now we are ready to establish the empirical analogue of Theorem 2. Note that the μ_k are random probability measures (depending on the path $\zeta \in \boldsymbol{X}^\infty$), and so the quantities we will deal with next are, in fact, random variables. Regarding the approximation of the optimal average value g^* of $\mathcal{M}$ for an initial state $x \in \boldsymbol{X}$, we proceed as in Section 3.2 with the empirical probability measures μ_k, so as to obtain the (random) approximation $\mathcal{J}^*_{k,\delta}(x)$.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Theorem 3. *Suppose that the control model $\mathcal{M}$ satisfies Assumptions 1–3 with $\mu \in \mathcal{P}_{\exp}(\boldsymbol{X})$. Fix an initial state $x \in \boldsymbol{X}$ and let $\varepsilon > 0$ be some given precision. There exist $\delta > 0$, and positive constants $C(\varepsilon)$ and $D(\varepsilon)$ such that*

$$\mathbb{P}_\mu\left\{\sup_{x\in\boldsymbol{X}} |g^* - \mathcal{J}^*_{k,\delta}(x)| > \varepsilon\right\} \leq C(\varepsilon)\mathrm{e}^{-D(\varepsilon)k} \quad \textit{for all } k \geq 1.$$

Proof. Consider $\varepsilon > 0$. Recalling Theorem 2, we have

$$\left\{\sup_{x\in\boldsymbol{X}} |\mathcal{J}^*(x) - \mathcal{J}^*_{k,\delta}(x)| > \varepsilon\right\} \subseteq \left\{\mathcal{W}(\mu, \mu_k) > \frac{\varepsilon}{2\mathcal{D}}\right\} \cup \left\{\mathcal{W}(\mu, \mu_k) > \frac{\rho}{2(L_{wp} + L_p)}\right\} \quad \text{for } \delta < \varepsilon/2\mathcal{C}.$$

By Proposition 3, there exist $C(\varepsilon)$ and $D(\varepsilon)$ such that

$$\mathbb{P}_\mu\left\{\mathcal{W}(\mu, \mu_k) > \left[\frac{\varepsilon}{2\mathcal{D}} \vee \frac{\rho}{2(L_{wp} + L_p)}\right]\right\} \leq C(\varepsilon)\mathrm{e}^{-D(\varepsilon)k} \quad \text{for all } k \geq 1,$$

yielding the result. □

This theorem states that, given some initial state $x \in \boldsymbol{X}$ and some precision $\varepsilon > 0$, we can find $\delta > 0$ sufficiently small such that when approximating g^* by $\mathcal{J}^*_{k,\delta}(x)$, the probability of not reaching the precision ε goes to 0 exponentially in the sample size k.

4.2. The deterministic approach

This approach consists in deriving the approximating probability measure μ_k from a covering of $\boldsymbol{X}$ with small radius. This allows us to tightly control the distance $\mathcal{W}(\mu, \mu_k)$ but it may pose an additional computational challenge. The next result is borrowed from [1, Proposition 1.1].

Proposition 4. *Consider the Borel state space $\boldsymbol{X}$. Given $\mu \in \mathcal{P}_1(\boldsymbol{X})$ and $\varepsilon > 0$, there exists $\nu \in \mathcal{P}_1(\boldsymbol{X})$ with finite support such that $\mathcal{W}(\mu, \nu) \leq \varepsilon$.*

Proof. Consider a measurable partition $\{B_n\}_{n\geq 1}$ of $\boldsymbol{X}$ such that, for each $n \geq 1$, there is some $y_n \in B_n$ with $\rho_{\boldsymbol{X}}(y, y_n) \leq \frac{1}{3}\varepsilon$ for all $y \in B_n$. Such a construction is possible because $\boldsymbol{X}$ is separable. The probability measure $\lambda = \sum_{n=1}^{\infty} b_n \delta_{y_n}$ with $b_n = \mu(B_n)$ verifies $\lambda \in \mathcal{P}_1(\boldsymbol{X})$ and $\mathcal{W}(\lambda, \mu) \leq \frac{1}{3}\varepsilon$. Choose N such that

$$\sum_{r>N} \int_{B_r} \rho_{\boldsymbol{X}}(y, y_1)\mu(\mathrm{d}y) \leq \tfrac{1}{3}\varepsilon, \tag{13}$$

which implies $\sum_{r>N} b_r \rho_{\boldsymbol{X}}(y_r, y_1) \leq \frac{2}{3}\varepsilon$, and define

$$\nu = \left(b_1 + \sum_{r>N} b_r\right)\delta_{y_1} + \sum_{k=2}^{N} b_k \delta_{y_k}.$$

We have $\mathcal{W}(\nu, \lambda) \leq \frac{2}{3}\varepsilon$ and, therefore, $\mathcal{W}(\nu, \mu) \leq \varepsilon$. □

Consequently, given some precision ε, we can construct a probability measure μ_k such that $\mathcal{W}(\mu, \mu_k) \leq \varepsilon$, where the number of points k in the support of μ_k is given by $k = N_\varepsilon$ needed to achieve (13).

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

5. Numerical example

Consider a queueing system with finite capacity. The state space is $\boldsymbol{X} = [0, C]$ for some $C > 0$ and the action space is some interval $\boldsymbol{A} = [a_m, a_M]$. We suppose that $\boldsymbol{A}(x) = \boldsymbol{A}$ for all $x \in \boldsymbol{X}$. The transition kernel is given by

$$q(B \mid x, a) = \int_{B \cap (x, C]} 2(y - x)\,\mathrm{d}x + a\,\mathbf{1}_B(0) - (a + (C - x)^2)\,\mathbf{1}_B(x)$$

for any $0 \le x \le C$, $a_m \le a \le a_M$, and $B \in \mathcal{B}(\boldsymbol{X})$. The cost rate function $c\colon \boldsymbol{X} \times \boldsymbol{A} \to$ is bounded and Lipschitz continuous.

Proposition 5. *Consider the queueing system defined above. If* $a_m > (1 + C)(2C + 1)$ *then all the assumptions in this paper are satisfied.*

Proof. In $\boldsymbol{X}$ we suppose that the point 0 is 'separated' from the interval $(0, C]$, and we will suppose that $d_{\boldsymbol{X}}$ is the usual metric on $(0, C]$ with $d_{\boldsymbol{X}}(0, x) = 1 + x$ for $0 < x \le C$. All the conditions in Assumptions 1 and 2 are easy to verify except Assumptions 1(iv) and 2(ii).

Regarding Assumption 2(ii) we have the following. Let $w \equiv 1$ be constant, and [10, Assumptions A, B, and C(i)] are satisfied. Define $B = \{0\}$; see [10, p. 961]. For all $x \ne 0$, we have $x \hookrightarrow B$ and the condition of [10, Theorem 3.3(c)] holds. Then we use [10, Theorem 3.3(d) and Equation (5.24)] to deduce that [10, Assumption C] holds with $\sup_{\alpha > 0} \|h_\alpha\| < \infty$ in the supremum norm. Hence, Assumption 2(ii) indeed holds.

Now we turn to Assumption 1(iv). In our case, we have $L_\Psi = 0$. For the function v in Assumption 1(iv), assume without loss of generality that $v(0) = 0$. Then $|v(x)| \le L_v(1 + C)$ for all $x \in \boldsymbol{X}$. We need to find the Lipschitz constant of

$$\frac{1}{\hat{q}} \int_x^C 2v(u)(u - x)\,\mathrm{d}u + v(x)\left(1 - \frac{(C - x)^2 + a}{\hat{q}}\right), \tag{14}$$

and show that there exists some $L_Q < 1$ such that this Lipschitz constant is less than $L_Q L_v$.

The Lipschitz constant of the first term in (14) is $L_v \boldsymbol{p}(C)/\hat{q}$, where $\boldsymbol{p}(C)$ is some polynomial in C. It can be made small by choosing large $\hat{q}$ (for the calculations, bound the derivative of this term and use $\|v\| \le L_v(1 + C)$; in this way, we obtain some constants multiplied by $L_v/\hat{q}$). For the second term, note that it is the product of two functions:

- v, which is L_v-Lipschitz continuous and bounded by $L_v(1 + C)$;
- $1 - ((C - x)^2 + a)/\hat{q}$, which is $(2C + 1)/\hat{q}$-Lipschitz continuous, and is nonnegative for large enough $\hat{q}$ and bounded by $1 - a_m/\hat{q}$.

Therefore, the product is

$$L_v\left(1 - \frac{a_m}{\hat{q}}\right) + L_v(1 + C)\frac{2C + 1}{\hat{q}}$$

Lipschitz continuous. The term that multiplies L_v is less than 1 since $a_m > (1 + C)(2C + 1)$.

In this proof, note that in (14) we must use a sharp bound for the second term because we need $L_Q < 1$. So we need the term $1 - ((C - x)^2 + a)/\hat{q}$ to be bounded away from 1.

Now we turn to Assumption 3. Fix arbitrary $0 < \eta < 1$ and define the probability measure μ on $\boldsymbol{X}$ as

$$\mu = \eta\delta_0 + \frac{1 - \eta}{C}\lambda,$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

where λ is the Lebesgue measure on $(0, C]$. Then, for any $(x, a) \in \boldsymbol{K}$, we have

$$p(y \mid x, a) = \frac{2C}{1-\eta}(y - x) \quad \text{if } 0 < y \leq C$$

and $p(0 \mid x, a) = a/\eta$. For $0 < \delta < 1$, we choose $\boldsymbol{A}_\delta(x)$ to be $[1/\delta] + 1$ equally spaced points in $\boldsymbol{A}$, namely,

$$\boldsymbol{A}_\delta = \left\{ a_m + \frac{j(a_M - a_m)}{[1/\delta]} \right\}_{j=0,1,\ldots,[1/\delta]} \tag{15}$$

with $d_{\mathrm{H}}(\boldsymbol{A}, \boldsymbol{A}_\delta) \leq \delta$ (we do not make the dependence on x explicit since the action sets $\boldsymbol{A}_\delta(x)$ are the same for all x). It is straightforward to now check that Assumption 3 holds. □

Given $k \geq 1$, define the probability measure μ_k as

$$\mu_k = \eta\delta_0 + \frac{1-\eta}{k} \sum_{j=1}^{k} \delta_{x_j}$$

with $x_1, \ldots, x_k \in (0, C]$.

(a) If we follow the empirical approach described in Section 4.1 then we interpret the $\{x_i\}$ as a sample of size k of a uniform law on $(0, C]$.

(b) By using the deterministic approach in Section 4.2, we can let $x_j = jC/k$ for $1 \leq j \leq k$, which yields

$$\mathcal{W}(\mu, \mu_k) = \frac{(1-\eta)C}{2k}.$$

The control model $\mathcal{M}_{k,\delta}$ is given by the following transition rates (we specify them only on the support Γ_k of μ_k). Given $x, y \in \Gamma_k$ with $x \neq y$, and $a \in \boldsymbol{A}_\delta$,

$$q_k(y \mid x, a) = \frac{2(y-x)^+ C}{k} \quad \text{and} \quad q_k(0 \mid x, a) = a.$$

The transition rate $q_k(x \mid x, a)$ is equal to $-\sum_{y \neq x} q_k(y \mid x, a)$. Note that the transition rates of $\mathcal{M}_{k,\delta}$ do not depend on η. We can solve each $\mathcal{M}_{k,\delta}$ by using the policy iteration algorithm, which converges in a finite number of steps.

5.1. Numerical experimentation

We choose $C = 1$, $\boldsymbol{A} = [7, 8]$, and $c(x, a) = (1 - x)(10 - a)$. For any $k \geq 1$, for the discretization of the state space, we choose k points in $(0, C]$ (details will be given later), while for the discretization of the action space we let $\delta = 1/k$ (recall (15)). Since the dependence is only on k, the approximating control model will be denoted by $\mathcal{M}_k$ in lieu of $\mathcal{M}_{k,\delta}$.

The empirical approach. Given a value of $k \geq 1$, we take 10 000 samples of size k of the uniform law in $(0, C]$. For each sample we solve the problem $\mathcal{M}_k$: this yields 10 000 observations of the constant (in x) random optimal cost g_k^*. In Table 1 we present the results for the mean $\overline{g_k^*}$ and the standard deviation of the 10 000 optimal costs for several values of k.

The estimations are very stable (the mean is practically the same for all values of k) and they become more concentrated as k grows (the standard deviation decreases). In Figure 1 we present the density estimators (based on normal kernels) of the 10 000 optimal costs for these values of k. We also observe that the estimations become more accurate and concentrated as k grows.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

TABLE 1: Estimation of the optimal average cost.

k	Mean	Standard deviation
30	1.8449	0.0243
60	1.8450	0.0172
90	1.8449	0.0139
120	1.8450	0.0122

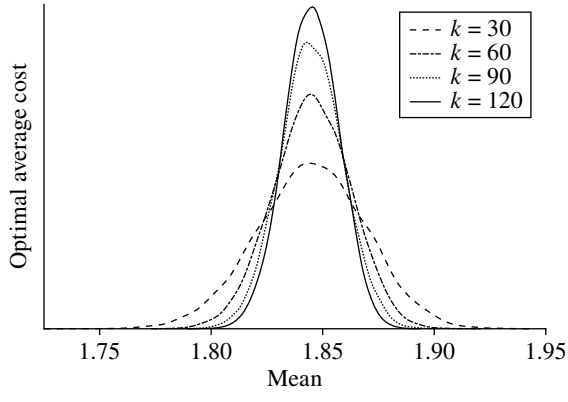

FIGURE 1: Density estimators for the estimation of the optimal average cost.

To check empirically the order of convergence given in Theorem 3, we fix $\varepsilon > 0.01$ and we estimate $\mathbb{P}\{|g^* - g_k^*| > \varepsilon\}$ by p_k, defined as the empirical probability for the 10 000 samples that $|g_k^* - \overline{g_k^*}| > 0.01$. Such calculations are made for $k = 10, 20, \ldots, 120$. We then perform a linear regression

$$-\log p_k \sim \beta_0 + \beta_1 k;$$

see Figure 2(a). The linear fit is very precise, with an R-squared coefficient of 0.983. We obtain $\hat{\beta}_0 = 0.1955$ and $\hat{\beta}_1 = 0.006$, which yields the empirical approximation

$$\mathbb{P}\{|g^* - g_k^*| > 0.01\} \simeq 0.8224 \times \mathrm{e}^{-0.006k}.$$

For $\varepsilon = 0.02$, we perform a similar linear regression analysis which exhibits again a very good linear fit (see Figure 2(b)), this time with an R-squared coefficient of 0.993 and an approximation

$$\mathbb{P}\{|g^* - g_k^*| > 0.02\} \simeq 0.682 \times \mathrm{e}^{-0.0165k}.$$

From this, we see (empirically) that the *nonasymptotic bound* given in Theorem 3 is tight (its order is indeed attained) and we are able to estimate the involved constants. The multiplicative constant $C(\varepsilon)$, in particular, takes a 'reasonably' low value.

The deterministic approach. We use now the deterministic approach described in (b) above for the probability measure μ_k. The nature of the discretized model $\mathcal{M}_k$ means that its optimal average cost g_k^* is constant, and Theorem 2 ensures that

$$|g^* - g_k^*| = O\left(\frac{1}{k}\right). \tag{16}$$

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

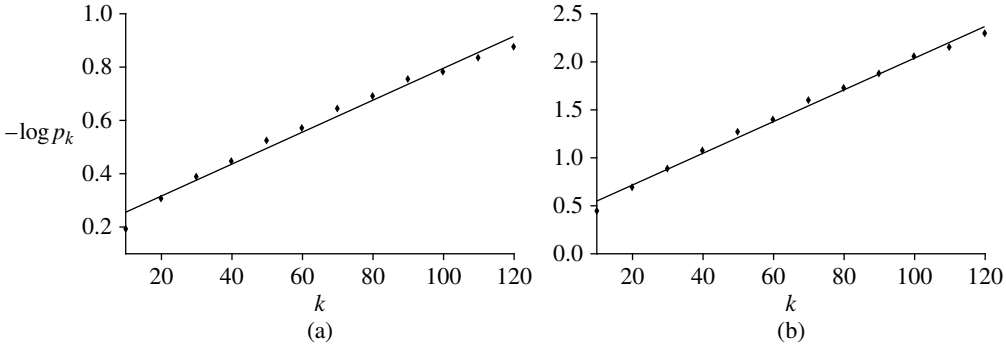

FIGURE 2: Regression plots of $-\log p_k$ with (a) error $\varepsilon = 0.01$ and (b) error $\varepsilon = 0.02$.

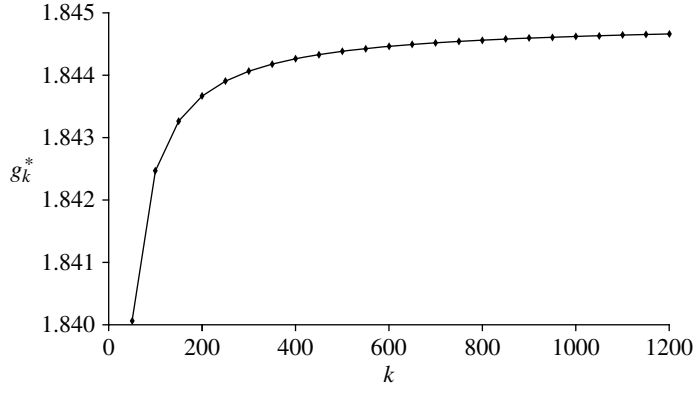

FIGURE 3: Optimal average cost g_k^* of $\mathcal{M}_k$.

For values of k ranging from 50 to 1200 (with a step size of 50), we have solved the approximating control model $\mathcal{M}_k$. In Figure 3 we display the corresponding optimal average cost g_k^*. We observe that the approximations g_k^* become very stable as k grows. To study the dependence of g_k^* on k, and in order to check empirically (16), we perform a linear regression analysis of the form

$$g_k^* \sim \beta_0 + \beta_1 \frac{1}{k}.$$

Using our 24 observations for $k = 50, 100, \ldots, 1200$, this yields the estimator $\hat{\beta}_0 = 1.8448$ for the optimal average cost g^*, and $\hat{\beta}_1 = -0.0048$ with residuals satisfying

$$\max_k \left| g_k^* - \hat{\beta}_0 - \frac{\hat{\beta}_1}{k} \right| \leq 4 \times 10^{-6}$$

with a squared correlation coefficient of $R^2 = 1$. So, the approximations g_k^* and the regression line $\hat{\beta}_0 + \hat{\beta}_1/k$ almost overlap. We see (empirically, for this example) that the *nonasymptotic bound* in (16) is tight (its order is indeed attained). For this example, the constant in the $O(1/k)$-term is fairly small, of order 5×10^{-3}.

https://doi.org/10.1017/jpr.2018.36 Published online by Cambridge University Press

Acknowledgement

Research supported by grant MTM2016-75497-P from the Spanish Ministerio de Economía y Competitividad.

References

[1] Anselmi, J., Dufour, F. and Prieto-Rumeau, T. (2016). Computable approximations for continuous-time Markov decision processes on Borel spaces based on empirical measures. *J. Math. Anal. Appl.* **443,** 1323–1361.
[2] Bertsekas, D. P. and Tsitsiklis, J. N. (1996). *Neuro-Dynamic Programming*. Athena Scientific, Belmont, MA.
[3] Boissard, E. (2011). Simple bounds for convergence of empirical and occupation measures in 1-Wasserstein distance. *Electron. J. Prob.* **16,** 2296–2333.
[4] Chang, H. S., Fu, M. C., Hu, J. and Marcus, S. I. (2007). *Simulation-Based Algorithms for Markov Decision Processes*. Springer, London.
[5] Dufour, F. and Prieto-Rumeau, T. (2012). Approximation of Markov decision processes with general state space. *J. Math. Anal. Appl.* **388,** 1254–1267.
[6] Dufour, F. and Prieto-Rumeau, T. (2013). Finite linear programming approximations of constrained discounted Markov decision processes. *SIAM J. Control Optimization* **51,** 1298–1324.
[7] Dufour, F. and Prieto-Rumeau, T. (2014). Stochastic approximations of constrained discounted Markov decision processes. *J. Math. Anal. Appl.* **413,** 856–879.
[8] Dufour, F. and Prieto-Rumeau, T. (2015). Approximation of average cost Markov decision processes using empirical distributions and concentration inequalities. *Stochastics* **87,** 273–307.
[9] Guo, X. and Rieder, U. (2006). Average optimality for continuous-time Markov decision processes in Polish spaces. *Ann. Appl. Prob.* **16,** 730–756.
[10] Guo, X. and Ye, L. (2010). New discount and average optimality conditions for continuous-time Markov decision processes. *Adv. Appl. Prob.* **42,** 953–985.
[11] Guo, X. and Zhang, W. (2014). Convergence of controlled models and finite-state approximation for discounted continuous-time Markov decision processes with constraints. *Europ. J. Operat. Res.* **238,** 486–496.
[12] Hinderer, K. (2005). Lipschitz continuity of value functions in Markovian decision processes. *Math. Meth. Operat. Res.* **62,** 3–22.
[13] Jacod, J. (1979). *Calcul Stochastique et Problèmes de Martingales* (Lecture Notes Math. **714**). Springer, Berlin.
[14] Piunovskiy, A. and Zhang, Y. (2014). Discounted continuous-time Markov decision processes with unbounded rates and randomized history-dependent policies: the dynamic programming approach. *4OR* **12,** 49–75.
[15] Powell, W. B. (2007). *Approximate Dynamic Programming*. John Wiley, Hoboken, NJ.
[16] Prieto-Rumeau, T. and Hernández-Lerma, O. (2012). Discounted continuous-time controlled Markov chains: convergence of control models. *J. Appl. Prob.* **49,** 1072–1090.
[17] Prieto-Rumeau, T. and Lorenzo, J. M. (2010). Approximating ergodic average reward continuous-time controlled Markov chains. *IEEE Trans. Automatic Control* **55,** 201–207.
[18] Saldi, N., Linder, T. and Yüksel, S. (2015). Asymptotic optimality and rates of convergence of quantized stationary policies in stochastic control. *IEEE Trans. Automatic Control* **60,** 553–558.
[19] Sutton, R. S. and Barto, A. G. (1998). *Reinforcement Learning: An Introduction*. MIT Press, Cambridge, MA.
[20] Van Roy, B. (2002). Neuro-dynamic programming: overview and recent trends. In *Handbook of Markov Decision Processes* (Internat. Ser. Operat. Res. Manag. Sci. **40**), Kluwer, Boston, MA, pp. 431–459.
[21] Zhang, Y. (2014). Average optimality for continuous-time Markov decision processes under weak continuity conditions. *J. Appl. Prob.* **51,** 954–970.